

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **04241400 A**

(43) Date of publication of application: **28.08.92**

(51) Int. Cl.

G10L 9/08
G06F 15/18

(21) Application number: **03003219**

(22) Date of filing: **16.01.91**

(71) Applicant: **OKI ELECTRIC IND CO LTD**

(72) Inventor: **NOTO HIROYUKI**

(54) FUNDAMENTAL FREQUENCY EXTRACTING METHOD

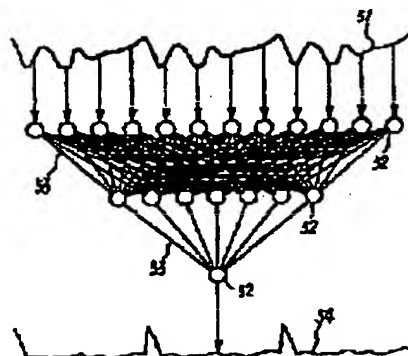
(57) Abstract:

PURPOSE: To accurately detect the time of the epoch point of a signal waveform and extract an accurate fundamental frequency by using a neural network which has a specific function constituted by a learning process.

CONSTITUTION: The perceptron type neural network which performs a nonlinear emphasizing process is so constituted that respective process units (cell) in the input layer are connected to cells 52 in the intermediate layer through a communication circuit 53 and the respective cells in the intermediate layer are connected to cells 52 in the output layer. A display signal for learning is inputted to the neural network and the learning process is carried out by an error reverse propagating method. The neural network generates a peak output only when a specific position in a time interval of the inputted display signal coincides with the epoch point. An input signal 51 is inputted to the neural network wherein the learning is performed as mentioned above and peaks of an output sequence 54 obtained by shifting the time of the input signal 51, bit by bit, are detected as epoch points. Then the fundamental frequency of the input signal 51 is

calculated from the time intervals of the respective epoch points.

COPYRIGHT: (C)1992,JPO&Japio



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平4-241400

(43)公開日 平成4年(1992)8月28日

(51)Int.Cl.⁹

識別記号

庁内整理番号

F I

技術表示箇所

G 1 0 L 9/08

B 8946-5H

G 0 6 F 15/18

8945-5L

審査請求 未請求 請求項の数2(全 8 頁)

(21)出願番号 特願平3-3219

(22)出願日 平成3年(1991)1月16日

(71)出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72)発明者 野戸 広之

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(74)代理人 弁理士 柿本 恭成

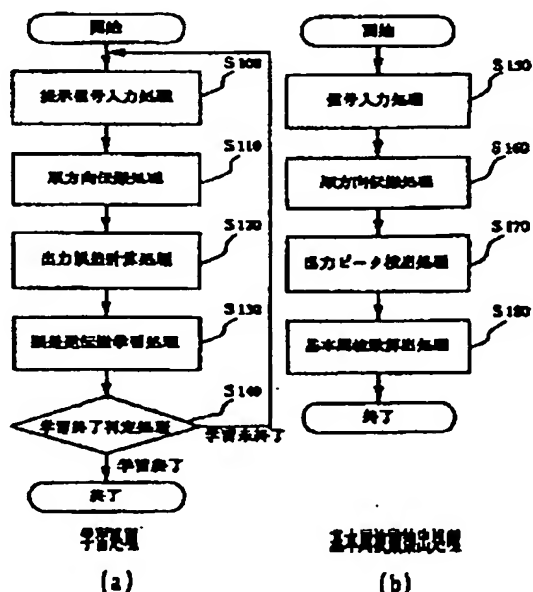
(54)【発明の名称】 基本周波数抽出方法

(57)【要約】

【目的】エポック点の検出を正確に行い、正確な基本周波数を抽出する方法を提供する。

【構成】パーセプトロン型の神経回路網に対し、ステップS100で学習のための提示信号の大きさを一定の時間区間にわたって正規化した信号値を入力し、ステップS110～S140で入力した時間区間の予め定められた位置と提示信号のエポック点とが一致したときのみ出力が最大を取り、誤差逆伝搬法による学習処理を施す。このように学習処理がなされた神経回路網を用い、ステップS150～S180で基本周波数が未知の入力信号の大きさを一定の時間区間にわたって正規化した信号値を前記神経回路網に入力し、入力信号の時刻をわずかつ移動させて得られる神経回路網の出力系列のピークをエポック点として検出し、該エポック点の時刻の間隔から基本周波数を計算する。

本発明の実施例



1

2

【特許請求の範囲】

【請求項1】 所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、該検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出することを特徴とする基本周波数抽出方法。

【請求項2】 請求項1記載の基本周波数抽出方法において、前記提示信号及び前記入力信号は、電力によって正規化された基本周波数抽出方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、音声信号、音響信号、及び画像信号等の信号の基本周波数を抽出する基本周波数抽出方法に関するものである。

【0002】

【従来の技術】 従来、この種の分野の技術としては、例えば図2のようなものがあった。以下、図2を参照しつつ従来の基本周波数抽出方法について説明する。図2は従来の基本周波数抽出方法の処理手順を示すフローチャートであり、各処理を順番にステップS1～S6で表す。まず、処理の対象となる信号波形に窓関数を乗じて所望の時間領域の信号を取り出した後（ステップS1）、信号波形の自己相関関数を求める（ステップ2）。続いて、ステップ2で求められた自己相関関数をもとに線形予測係数を求め（ステップS3）、さらに、この線形予測係数に基づき信号波形に対する線形予測を行って予測残差信号を求める（ステップS4）。その後、予測残差信号のピークを抽出してエポック点を検出し（ステップ5）、エポック点の時間間隔から基本周波数を計算する（ステップ6）。ここで、エポック点と*

*は、信号波形の基本周波数に対応した時刻であり、音声信号の場合は、例えば声帯が最も収縮した発声時の時刻をいう。

【0003】 次に、各処理の内容を説明する。

（1）ステップS1の処理

時間領域で離散的な信号波形を $s(m)$ とし（但し、 m は離散的な時刻）、適当な窓関数を $w(m)$ とする。今、処理対象となる信号波形のうち、基本周波数を求めたい所望の離散的な時刻を n とする。このとき、所望の時刻 n における信号波形 $s_n(m)$ は次式で求める。

【0004】

【数1】

【0005】 $s_n(m) = s(m+n)w(m)$

但し、 $0 \leq m \leq N-1$ 、（ N ：所望の窓関数の大きさ）

（2）ステップS2の処理

所望の時刻 n における信号波形の自己相関関数 $R_n(k)$ を次式により求める。

【0006】

【数2】

20 【0007】

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k)$$

但し、 k ：ラグ

（3）ステップS3の処理

線形予測係数を求める。ダービン（Dublin）の再帰法によれば、次式によって線形予測係数 α_j を求めることができる。ここで、 j は第 j 次の予測係数であることを示し、 p は線形予測の次数で任意である。

30 【0008】

【数3】

【0009】 $E^{(0)} = R(0)$

【0010】

【数4】

【0011】

$$k_i = (R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)) / E^{(i-1)} \quad 1 \leq i \leq p$$

【0012】

【数5】

【0013】 $\alpha_i^{(1)} = k_i$

40 ※ 【0014】

【数6】

※ 【0015】

$$\alpha_i^{(j)} = \alpha_i^{(j-1)} - k_i \alpha_{i-1}^{(j-1)} \quad 1 \leq j \leq i-1$$

【0016】

【数8】

【0019】 $\alpha_i = \alpha_i^{(p)} \quad 1 \leq i \leq p$

として線形予測係数 α_i を得る。

【0020】 （4）ステップS4の処理

ステップS3で求めた線形予測係数 α_i を用いて予測残差信号 $e(n)$ を求める。予測残差信号 $e(n)$ は次式で計算される。

50

計算の手順としては、まず、数3により $E^{(0)}$ を得る。次に、数4から数7までの計算を行い、 $\alpha_i^{(1)}$ を $1 \leq i \leq p$ の範囲で順に再帰的に求める。そして、最後に、

【0018】

【0021】

【数9】

*【0022】

*

$$e(n) = s(n) - \sum_{i=1}^n \alpha_i s(n-i)$$

(5) ステップS5の処理

残差信号 $e(n)$ のピークをエポック点として求める。この処理はある一定の閾値 H を定め、残差信号 $e(n)$ がこの閾値 H を終えてピークを生じた離散的な時刻をエポック点とし、各エポック点の離散的な時刻を T_0, T_1, T_2, \dots とする。

(6) ステップS6の処理

最後にエポック点の間隔から b 番目のピッチにおける基本周波数 f_b を次式により計算する。

【0023】

【数10】

【0024】 $f_b = f_1 / (T_b - T_{b-1})$

【0025】

【発明が解決しようとする課題】しかしながら、上記の基本周波数抽出方法では、次のような課題があった。① 予測残差信号 $e(n)$ の波高値の変化範囲は一定ではないため、エポック点検出のための閾値 H の設定によってはエポック点が正確に検出できない。

【0026】② 予測残差信号 $e(n)$ の波形は複雑であり、ピークが多数生ずる。そのため、どのピークがエポック点によるものかが判別しにくく、正確な基本周波数の検出が困難である。本発明は前記従来技術の持っていた課題として、正確なエポック点の検出が困難である点、正確な基本周波数の抽出が困難である点について解決した基本周波数抽出方法を提供するものである。

【0027】

【課題を解決するための手段】本発明は、前記課題を解決するために、所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、該検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出するようにしたものである。また、前記提示信号及び前記入力信号は、電力によって正規化してもよい。

【0028】

【作用】本発明は、以上のように基本周波数抽出方法を構成したので、神経回路網に提示信号を入力させて誤差逆伝搬法により学習処理を施せば、神経回路網は、入力された提示信号の時間区間の所定の位置とエポック点との一致時のみに出力がピークとなるような信号波形を出

力して、基本周波数を有する信号が入力されるとその信号のエポック点が自動的に検出されるようになる。このように学習がなされた神経回路網に基本周波数が未知である入力信号を入力し、その入力信号の時刻をわずかつ移動させて得られる神経回路網の出力系列のピークをエポック点として複数、検出した後、そのエポック点の時間間隔から入力信号の基本周波数を計算する。このように、非線形処理によってエポック点の時刻を強調して出力することにより、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出できる。したがって、前記課題を解決できるのである。

【0029】

【実施例】図1(a)、(b)は、本発明の実施例を示す基本周波数抽出方法の処理手順を示すフローチャートであり、同図(a)は学習処理の手順を示すフローチャート及び同図(b)は基本周波数抽出処理の手順を示すフローチャートである。また、図3は本発明の実施例を示す基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。図3において、この基本周波数抽出装置は、処理対象となる音声信号等の信号波形51をアナログ/ディジタル変換してディジタルデータを求める図示しないアナログ/ディジタル変換器を有し、その出力側には入力層を形成する複数個の処理ユニット(以下、単にセルという)52が接続されている。さらに、通信回路53を介して入力層の各セル52が中間層のセル52に、中間層の各セル52が出力層のセル52にそれぞれ接続されている。これら各セル52は神経細胞に相当するものであり、それぞれマイクロプロセッサで構成されている。通信回路53は神経細胞間の接続ニューロンに対応するものであり、ローカルエリアネットワークで構成されている。さらに、出力層のセル52から出力される出力系列54をディジタルデータとして蓄積する図示しないメモリが設けられている。このようなハード構成により、非線形強調処理を行うパーセプトロン型の神経回路網が形成される。

【0030】この神経回路網の処理内容を以下に説明する。本発明の処理は、神経回路網の学習処理とその学習処理が終わった神経回路網による基本周波数の抽出処理とに別れる。まず、神経回路網の学習処理について図1(a)及び図4を参照しつつ説明する。なお、図4は学習処理における各信号の説明図である。神経回路網の学習のために入力として加える提示信号300として、男性の発生した母音波形を12kHz、12ビットでサンプリングしたものをを用い、さらに、提示信号300に対して人間の視察により教師信号のピークを与えるエポッ

5

ク点の位置310（以下、教師エポック点という）を予め設定しておく。また、時間領域で離散的な信号波形 $s(m)$ とし、提示信号300を特に $s_a(m)$ とする。

【0031】（1）提示信号入力処理（ステップS100）

提示信号300を入力するための処理は次のようにして行われる。まず、提示信号300を教師エポック点を中心とした区間から取り出し、教師信号として 0.9 を定める。この場合の提示信号300の時間区間330の長さは神経回路網の入力層のセル数に等しいサンプル数とし、本実施例では512点のサンプルを提示信号300とする。この場合の提示信号300を $s_{ao,j}(m)$ で表す。但し、 $0 \leq m \leq 511$ とする。そして、神経回路網の入力として、この提示信号 $s_{ao,j}(m)$ を次式によってパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{ij}^{(0)}$ とする。なお、 $o_{pj}^{(q)}$ は q 番目の層における p 番目のボタンに対する j 番目のセルの出力を表し、ここでは、提示信号 $s_{ao,j}(m)$ を1番目のボタン*

$$o_{ij}^{(0)} = 1 / (1 + \exp(-(\sum_{l=0}^{N_0-1} w_{li}^{(0)} o_{pj}^{(l-1)} + v_i^{(0)})))$$

ここで、 $1 \leq q \leq 2$ である。数12中の $o_{pj}^{(q)}$ は第 q 番目の層における第 j 番目のセルの出力であり、第 p 番目のボタンを提示した場合のものである。 N_0 は、第 q 番目の層におけるセルの数であり、 $w_{li}^{(q)}$ は第 $q-1$ 番目の層の第 i 番目のセルから第 q 番目の層の第 j 番目のセルへの重み係数、 $v_i^{(q)}$ は第 q 番目の層の第 j 番目のセルのバイアスである。本実施例では N_0 は512、 N_1 は64、 N_2 は1である。重み係数 $w_{li}^{(q)}$ とバイアス $v_i^{(q)}$ は学習前にはランダムな小さな値に設定しておく。数12の計算をすべての q に対して計算し、第2番目の出力層のセルの出力 $o_{pj}^{(2)}$ を得る。

【0036】（3）神経回路網の出力誤差計算処理（ステップS120）

q 番目の層の第 j 番目のセルにおける第 p 番目の入力に対する誤差を $\delta_{pj}^{(q)}$ で表すと、第 q 番目の提示信号に※

$$\delta_{pj}^{(q-1)} = o_{pj}^{(q-1)} (1 - o_{pj}^{(q-1)}) \sum_{k=1}^{N_{q-1}} \delta_{pk}^{(q)} w_{kj}^{(q)}$$

さらに誤差 $\delta_{pj}^{(q)}$ を用いて第 $(q-1)$ 層から第 q 層への重み係数 $w_{li}^{(q)}$ の修正量 $\Delta p w_{li}^{(q)}$ を次式で計算する。

【0041】

【数15】

【0042】

$$\Delta p w_{li}^{(q)} = \eta \delta_{pj}^{(q)} o_{pj}^{(q-1)}$$

また、第 q 層の第 j 番目のセルのバイアス $v_i^{(q)}$ に対する修正量 $\Delta p v_i^{(q)}$ も次式で計算する。

【0043】

【数16】

6

*とし、入力層を0番目の層としている。Cは正規化のための正定数である。

【0032】

【数11】

$$o_{ij}^{(0)} = \frac{s_{ao,j}(j)}{C \sqrt{\frac{\sum_{m=0}^{511} s_{ao,j}(m)^2}{512}}} + 0.5$$

10 【0033】（2）神経回路網の順方向伝搬処理（ステップS110）

本実施例の神経回路網の構造は入力層を第0番目の層として、第1番目の層を中間層、第2番目の層を出力層とする3層構造であり、第 q 番目の層の出力は次式で計算する。

【0034】

【数12】

【0035】

※対する教師信号を t_{pj} として、出力層における誤差 $\delta_{pj}^{(2)}$ を次式で計算する。

【0037】

【数13】

$$\delta_{pj}^{(2)} = (t_{pj} - o_{pj}^{(2)}) o_{pj}^{(2)} (1 - o_{pj}^{(2)})$$

（4）神経回路網の誤差逆伝搬学習処理（ステップS130）

30 第 q 番目の層の各セルの誤差 $\delta_{pj}^{(q)}$ が計算済みのとき、第 $(q-1)$ 層の各セルの誤差 $\delta_{pj}^{(q-1)}$ を次式で計算しておく。

【0039】

【数14】

【0040】

14-1

$$\delta_{pj}^{(q-1)} = o_{pj}^{(q-1)} (1 - o_{pj}^{(q-1)}) \sum_{k=1}^{N_{q-1}} \delta_{pk}^{(q)} w_{kj}^{(q)}$$

40 【0044】 $\Delta p w_{li}^{(q)} = \eta \delta_{pj}^{(q)}$

ここで、 η は学習の速度を決定する定数である。これらの計算を層の番号 q を減じながら全ての出力層、中間層について実行し、全ての重み係数 $w_{li}^{(q)}$ 及びバイアスに対する修正量 $\Delta p w_{li}^{(q)}$ 及び $\Delta p v_i^{(q)}$ を求め、全ての修正量 $\Delta p w_{li}^{(q)}$ 及び $\Delta p v_i^{(q)}$ を計算した後、この修正量を用いて全ての重み係数 $w_{li}^{(q)}$ 及びバイアス $v_i^{(q)}$ を次式によって修正する。

【0045】

【数17】

$$50 \quad w_{li}^{(q)} = w_{li}^{(q)} + \Delta p w_{li}^{(q)}$$

7

以上の処理を $p=1$ として、提示信号 $s_{a0.0}(m)$ に対して行う。この時、教師信号 t_{10} としては 0.9 を与える。

【0047】(5) 学習終了判定処理 (ステップ S140)

学習処理の繰返しによって重み係数 $w_{ji}^{(q)}$ が最適な値に近づくと、出力誤差 $\delta_{0.0}^{(2)}$ が 0 に近づく。出力誤差 $\delta_{0.0}^{(2)}$ が十分に小さな値 ε よりも小さな値になったかどうかを判定し、出力誤差 $\delta_{0.0}^{(2)}$ が大きければ学習未終了であると判定し、処理ステップ S100 に戻る。出力誤差 $\delta_{0.0}^{(2)}$ が小さければ、すべての学習処理を終了する。

【0048】次に、処理ステップ S100 に戻った場合には、提示信号として前述の教師エポック点を中心としない区間 320 をとる。区間 320 の中心の教師エポック点からのずれはランダムとする。この場合の提示信号 $s_{a0.1}(m)$ を次式でパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{2j}^{(0)}$ とする。

【0049】

【数18】

$$o_{2j}^{(0)} = \frac{s_{a0.1}(j)}{C \sqrt{\frac{\sum_{m=0}^{511} s_{a0.1}(m)^2}{512}}} + 0.5$$

【0050】このときの教師信号 t_{20} は 0.1 とし、処理ステップ S110 以降は前述と同様の処理を行う。以上の学習処理を繰返し行うことにより、最終的に最適な重み係数が得られる。

【0051】次に、学習後の神経回路網を用いた基本周波数抽出処理について、図1(b)及び図5を参照しつつ説明する。なお、図5は基本周波数抽出処理における各信号の説明図であり、符号400は基本周波数抽出の対象となる入力信号、符号420は神経回路網に入力するある1つの入力信号の区間、符号421は区間420に対する神経回路網の出力、符号430は神経回路網に入力する次の1つの入力信号の区間、符号431は区間430に対する神経回路網の出力、符号440は基本周波数抽出処理によって得られる神経回路網の出力系列、及び符号450は出力系列から抽出されたエポック点を示す。

【0052】(A) 神経回路網の信号入力処理 (ステップ S150)

時間領域で離散的な基本周波数が未知の入力信号を $x(m)$ とし、本実施例では学習処理に用いた提示信号とは別の話者の母音音声を入力信号として用いることにする。ここで、現在着目している時刻 u を中心とした時間区間における入力信号を $x_i^{(0)}$ とする。この入力信号を神経回路網の入力として次式によってパワ正規化して

8

オフセットを加え、入力層の各セルの出力 $o_{1j}^{(0)}$ とする。ここで、 $o_{1j}^{(0)}$ は q 番目の層における時刻 u を中心とした入力信号に対する j 番目のセルの出力を示し、 C は正規化のための正定数である。

【0053】

【数19】

$$o_{1j}^{(0)} = \frac{x_i(j)}{C \sqrt{\frac{\sum_{m=0}^{511} x_i(m)^2}{512}}} + 0.5$$

【0054】(B) 神経回路網の順方向伝搬処理 (ステップ S160)

この順方向伝搬処理は、数12における p を u に置き換えて計算することにより、学習処理における順方向伝搬処理と同様に行う。この処理によって出力層のセルからは出力 $o_{0.0}^{(2)}$ が得られる。さらに、入力信号を時刻 $u+1$ を中心とする時間区間からとり、同様の処理を行う。この様な処理を繰返し、時刻 u に対する出力 $o_{0.0}^{(2)}$ の系列を得る。この出力系列の一例を図5の符号440として示す。入力信号を取り出した時間区間の中心と入力信号のエポック点とが一致するすると、出力系列にはピークが生ずる。このピークを検出し、ピーク間の間隔を測定することによって入力信号の基本周波数を求めることができる。

【0055】(C) 神経回路網の出力ピーク検出処理 (ステップ S170)

出力系列に対するピーク検出処理は、出力が次式の条件を満たす離散的な時刻 v_d をエポック点の時刻として検出する。

【0056】

【数20】

【0057】 $o_{0.0}^{(2)}(v_d) \leq o_{0.0}^{(2)}(v_{d-1}) \geq o_{0.0}^{(2)}(v_{d+1})$, $P \leq o_{0.0}^{(2)}(v_d)$ ここで、 P はピークを検出するための閾値であり、本実施例では定数 0.5 を用いる。 d は検出したピークに付与する番号である。

【0058】(D) 基本周波数算出処理 (ステップ S180)

検出したエポック点の時刻 v_d の間隔から d 番目のピッチにおける基本周波数 f_d を次式により計算する。

【0059】

【数21】

【0060】 $f_d = f_s / (v_d - v_{d-1})$

本実施例は、信号波形のエポック点を自動的に検出する神経回路網を学習処理によって構成し、非線形処理によってエポック点の時刻を強調して出力するので、従来の線形予測法による残差信号を用いた基本周波数抽出方法に比べて以下のような利点がある。

(1) 神経回路網の出力セルから得られる出力 $o_{0.0}^{(2)}$

10

20

30

40

50

は、その値が0.0から1.0までの間であり、閾値 p は厳密な設定が不要である。

(2) 神経回路網の出力セルから得られる出力 $o_{ij}^{(2)}$ は、単純なパルス列となっており、ピーク位置の検出が容易であり、正確な基本周波数が検出しやすい。以上の利点を明確に示すために、実際のデータによって得られた各信号の例を図6(a)、(b)、(c)に示す。同図(a)は男性が発生した「ア」の一部である入力信号 $s(m)$ の波形図、同図(b)は線形予測分析によって得られた予測残差信号 $e(n)$ の波形図、及び同図(c)は神経回路網によって得られた出力系列 $o_{ij}^{(2)}$ の波形図である。なお、図中の縦線は人間の視察によって得られたエポック点の時刻であり、離散的な時刻 m 、 n 、 u の位置が互いに同一の時刻の部分を示している。図6(b)で明らかのように線形予測分析による残差信号波形 $e(n)$ は複雑であり、この信号波形からエポック点を正確に抽出して基本周波数を計算することは難しい。一方、神経回路網による出力系列 $o_{ij}^{(2)}$ は学習処理によって生成された最適な非線形処理により単純なインパルス列の形状を示し、この出力系列からエポック点を正確に抽出して基本周波数を計算することは容易である。

【0061】なお、本発明は、図示の実施例に限定されず、種々の変形が可能である。例えば、図4の提示信号300としては実施例に用いた以外の他の母音波形などの任意の周期性信号を与え、音響信号、画像信号等の一般的な信号波形に対してエポック点を検出できる神経回路網を構成することも可能である。

【0062】

【発明の効果】以上詳細に説明したように、本発明によれば、信号波形のエポック点を自動的に検出する神経回路網を学習処理によって構成し、その神経回路網に所定の時間区間にわたって正規化された入力信号を入力する。さらに、入力信号の時刻を移動させて得られる前記

神経回路網の出力系列のピークをエポック点として検出し、検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出するようにした。そのため、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出することが可能となる。また、提示信号及び入力信号を電力によって正規化すれば、正規化処理が簡単かつ的確に行える。

【図面の簡単な説明】

【図1】本発明の実施例を示す基本周波数抽出方法のフローチャートであり、同図(a)は学習処理のフローチャート及び同図(b)は基本周波数抽出処理のフローチャートである。

【図2】図2は従来の基本周波数抽出方法のフローチャートである。

【図3】本発明の実施例の基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。

【図4】本発明の実施例の学習処理における各信号の説明図である。

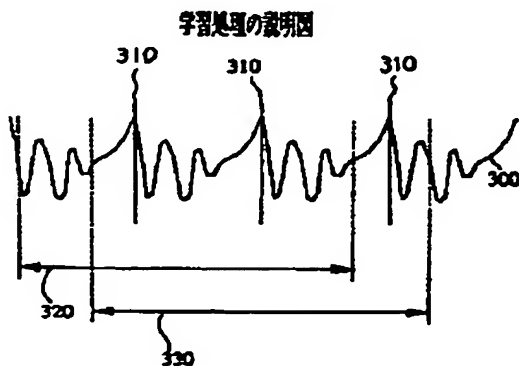
【図5】本発明の実施例の基本周波数抽出処理における各信号の説明図である。

【図6】本発明の効果を示す図であり、同図(a)は入力信号の波形図、同図(b)は従来の線形予測分析による波形図、及び同図(c)は神経回路網による出力系列の波形図である。

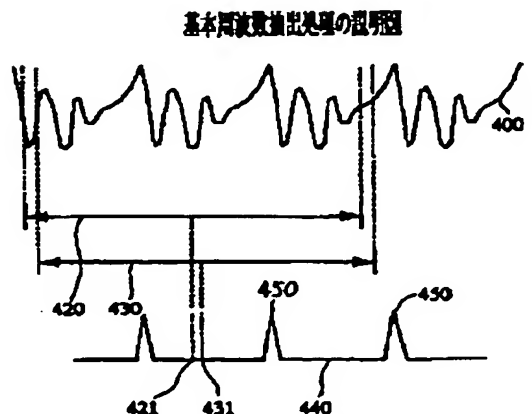
【符号の説明】

- 51 信号波形
- 52 セル
- 53 通信回線
- 54、440 出力系列
- 300 提示信号
- 310、450 エポック点
- 320、330、420、430 時間区間
- 400 入力信号

【図4】

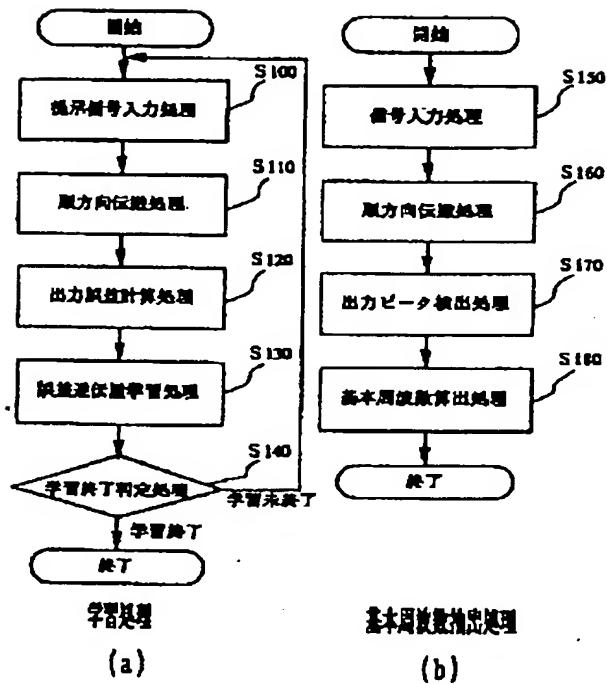


【図5】



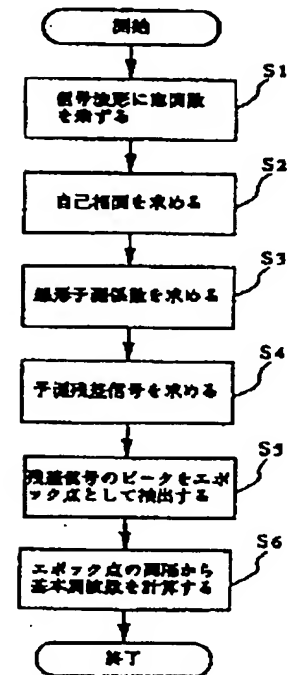
【図1】

本発明の実施例



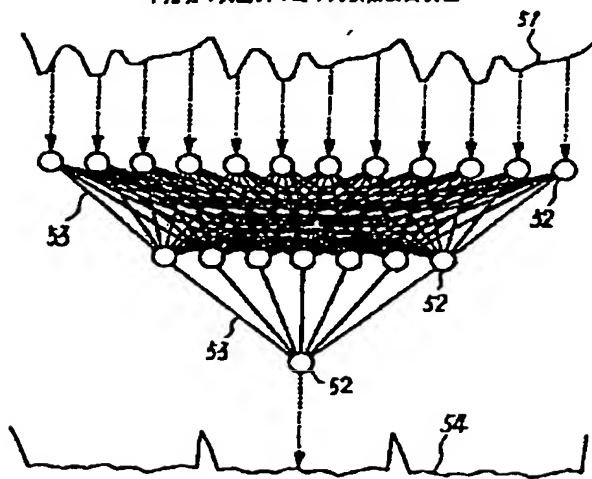
【図2】

従来の基本周波数抽出方法



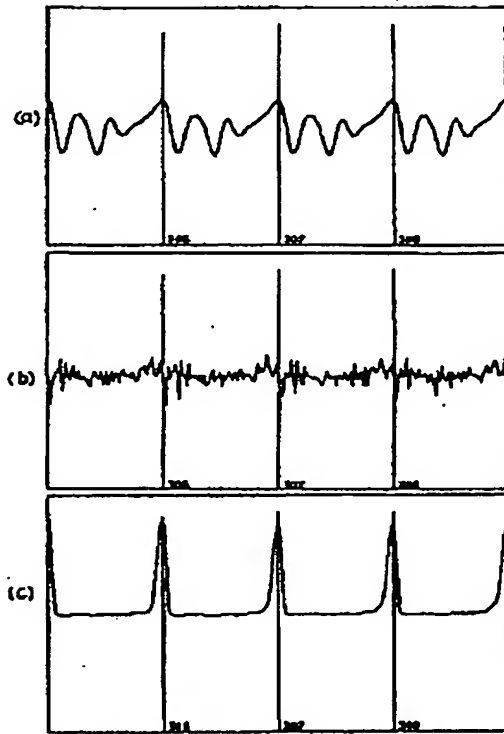
【図3】

本発明の実施例の基本周波数抽出装置



【図6】

本発明の差長を示す図



【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第2区分

【発行日】平成11年(1999)9月17日

【公開番号】特開平4-241400

【公開日】平成4年(1992)8月28日

【年通号数】公開特許公報4-2414

【出願番号】特願平3-3219

【国際特許分類第6版】

G10L 9/08

G06F 15/18

【FI】

G10L 9/08 B

G06F 15/18

【手続補正書】

【提出日】平成9年8月4日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】基本周波数抽出方法

【特許請求の範囲】

【請求項1】 所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、
所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、
前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、
該検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出することを特徴とする基本周波数抽出方法。

【請求項2】 請求項1記載の基本周波数抽出方法において、
前記提示信号及び前記入力信号は、電力によって正規化した基本周波数抽出方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声信号、音響信号、

$$s_n(m) = s(m+n)w(m)$$

但し、 $0 \leq m \leq N-1$ 、(N ；所望の窓関数の大きさ)

(2) ステップS2の処理

所望の時刻 n における信号波形の自己相関関数 R

及び画像信号等の信号の基本周波数を抽出する基本周波数抽出方法に関するものである。

【0002】

【従来の技術】従来、この種の分野の技術としては、例えば図2のようなものがあった。以下、図2を参照しつつ従来の基本周波数抽出方法について説明する。図2は従来の基本周波数抽出方法の処理手順を示すフローチャートであり、各処理を順番にステップS1～S6で表す。まず、処理の対象となる信号波形に窓関数を乗じて所望の時間領域の信号を取り出した後(ステップS1)、信号波形の自己相関関数を求める(ステップS2)。続いて、ステップS2で求められた自己相関関数をもとに線形予測係数を求め(ステップS3)、さらに、この線形予測係数に基づき信号波形に対する線形予測を行って予測残差信号を求める(ステップS4)。その後、予測残差信号のピークを抽出してエポック点を検出し(ステップS5)、エポック点の時間間隔から基本周波数を計算する(ステップS6)。ここで、エポック点とは、信号波形の基本周波数に対応した時刻であり、音声信号の場合は、例えば声帯が最も収縮した発声時の時刻をいう。

【0003】次に、各処理の内容を説明する。

(1) ステップS1の処理

時間領域で離散的な信号波形を $s(m)$ とし(但し、 m は離散的な時刻)、適当な窓関数を $w(m)$ とする。今、処理対象となる信号波形のうち、基本周波数を求めたい所望の離散的な時刻を n とする。このとき、所望の時刻 n における信号波形 $s_n(m)$ は次式(1)で求める。

$$\cdots (1)$$

$n(k)$ を次式(2)により求める。

【数1】

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k) \quad \dots (2)$$

但し、 k ; ラグ

【0004】 (3) ステップS3の処理
線形予測係数を求める。ダービン (Durbin) の再帰法によれば、次式 (3) ~ (7) によって線形予測係数 $E^{(0)} = R(0)$

数 α_j を求めることができる。ここで、 j は第 j 次の予測係数であることを示し、 p は線形予測の次数で任意である。

$\dots (3)$

【数2】

$$k_i = (R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)) / E^{(i-1)} \quad \dots (4)$$

但し、 $1 \leq i \leq p$

$$\alpha_j^{(i)} = k_i \quad \dots (5)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad \dots (6)$$

但し、 $1 \leq j \leq i-1$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad \dots (7)$$

計算の手順としては、まず、式 (3) により $E^{(0)}$ を得る。次に、式 (4) から式 (7) までの計算を行い、 $\alpha_j = \alpha_j^{(p)}$

$j^{(i)}$ を $1 \leq i \leq p$ の範囲で順に再帰的に求める。そして、最後に、

$\dots (8)$

但し、 $1 \leq j \leq p$

として線形予測係数 α_j を得る。

【0005】 (4) ステップS4の処理
ステップS3で求めた線形予測係数 α_j を用いて予測残差信号 $e(n)$

を求め

る。予測残差信号 $e(n)$ は次式 (9) で計算される。

【数3】

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad \dots (9)$$

(5) ステップS5の処理

残差信号 $e(n)$ のピークをエポック点として求める。この処理はある一定の閾値 H を定め、残差信号 $e(n)$ がこの閾値 H を終えてピークを生じた離散的な時刻をエポック点とし、各エポック点の離散的な時刻を T_0, T

T_1, T_2, \dots とする。

(6) ステップS6の処理

最後にエポック点の間隔とサンプリング周波数 f_s から、 b 番目のピッチにおける基本周波数 f_b を次式 (10) により計算する。

$$f_b = f_s / (T_b - T_{b-1}) \quad \dots (10)$$

法を提供するものである。

【0006】

【発明が解決しようとする課題】 しかしながら、上記の基本周波数抽出方法では、次のような課題があった。

① 予測残差信号 $e(n)$ の波高値の変化範囲は一定ではないため、エポック点検出のための閾値 H の設定によってはエポック点が正確に検出できない。

② 予測残差信号 $e(n)$ の波形は複雑であり、ピークが多数生ずる。そのため、どのピークがエポック点によるものが判別しにくく、正確な基本周波数の検出が困難である。

本発明は前記従来技術の持っていた課題として、正確なエポック点の検出が困難である点、正確な基本周波数の抽出が困難である点について解決した基本周波数抽出方

【0007】

【課題を解決するための手段】 本発明は、前記課題を解決するために、所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、該検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出するようにしたものである。また、前記提示信号及び前記入力信号は、電力によって正規化して

もよい。

【0008】

【作用】本発明は、以上のように基本周波数抽出方法を構成したので、神経回路網に提示信号を入力させて誤差逆伝搬法により学習処理を施せば、神経回路網は、入力された提示信号の時間区間の所定の位置とエポック点との一致時のみに出力がピークとなるような信号波形を出力して、基本周波数を有する信号が入力されるとその信号のエポック点が自動的に検出されるようになる。このように学習がなされた神経回路網に基本周波数が未知である入力信号を入力し、その入力信号の時刻をわずかつ移動させて得られる神経回路網の出力系列のピークをエポック点として複数、検出した後、そのエポック点の時間間隔から入力信号の基本周波数を計算する。このように、非線形処理によってエポック点の時刻を強調して出力することにより、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出できる。したがって、前記課題を解決できるのである。

【0009】

【実施例】図1(a)、(b)は、本発明の実施例を示す基本周波数抽出方法の処理手順を示すフローチャートであり、同図(a)は学習処理の手順を示すフローチャート及び同図(b)は基本周波数抽出処理の手順を示すフローチャートである。また、図3は本発明の実施例を示す基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。図3において、この基本周波数抽出装置は、処理対象となる音声信号等の信号波形51をアナログ／デジタル変換してデジタルデータを求める図示しないアナログ／デジタル変換器を有し、その出力側には入力層を形成する複数の処理ユニット(以下、単にセルという)52が接続されている。さらに、通信回路53を介して入力層の各セル52が中間層のセル52に、中間層の各セル52が出力層のセル52にそれぞれ接続されている。これら各セル52は神経細胞に相当するものであり、それぞれマイクロプロセッサで構成されている。通信回路53は神経細胞間の接続ニューロンに対応するものであり、ローカルエリアネットワークで構成されている。さらに、出力層のセル5

$$o_{ij}^{(0)} = \frac{S_{a0.9}(j)}{C \sqrt{\frac{\sum_{m=0}^{511} S_{a0.9}(m)^2}{512}}} + 0.5$$

(2) 神経回路網の順方向伝搬処理(ステップS110)

本実施例の神経回路網の構造は入力層を第0番目の層として、第1番目の層を中間層、第2番目の層を出力層と

2から出力される出力系列54をデジタルデータとして蓄積する図示しないメモリが設けられている。このようなハード構成により、非線形強調処理を行うパーセプトロン型の神経回路網が形成される。

【0010】この神経回路網の処理内容を以下に説明する。本発明の処理は、神経回路網の学習処理とその学習処理が終わった神経回路網による基本周波数の抽出処理とに分かれる。まず、神経回路網の学習処理について図1(a)及び図4を参照しつつ説明する。なお、図4は学習処理における各信号の説明図である。神経回路網の学習のために入力として加える提示信号300として、男性の発生した母音波形を12kHz、12ビットでサンプリングしたものをを用い、さらに、提示信号300に対して人間の視察により教師信号のピークを与えるエポック点の位置310(以下、教師エポック点という)を予め設定しておく。また、時間領域で離散的な信号波形s(m)とし、提示信号300を特にs_a(m)とする。

【0011】(1) 提示信号入力処理(ステップS100)

提示信号300を入力するための処理は次のようにして行われる。まず、提示信号300を教師エポック点を中心とした区間から取り出し、教師信号として0:9を定める。この場合の提示信号300の時間区間330の長さは神経回路網の入力層のセル数に等しいサンプル数とし、本実施例では512点のサンプルを提示信号300とする。この場合の提示信号300をs_{a0.9}(m)で表す。但し、0 ≤ m ≤ 511とする。そして、神経回路網の入力として、この提示信号s_{a0.9}(m)を次式(11)によってパワ正規化してオフセットを加え、入力層の各セルの出力o_{ij}⁽⁰⁾とする。なお、o_{pj}^(q)はq番目の層におけるp番目のボタンに対するj番目のセルの出力を表し、ここでは、提示信号s_{a0.9}(m)を1番目のボタンとし、入力層を0番目の層としている。Cは正規化のための正定数である。

【0012】

【数4】

$$\dots (11)$$

する3層構造であり、第q番目の層の出力は次式(12)で計算する。

【数5】

$$o_{pj}^{(q)} = 1 / (1 + \exp(-(\sum_{i=0}^{N_q-1} w_{ji}^{(q)} o_{pi}^{(q-1)} + v_j^{(q)}))) \quad \dots (12)$$

ここで、 $1 \leq q \leq 2$ である。式(12)中の $o_{pj}^{(q)}$ は第 q 番目の層における第 j 番目のセルの出力であり、第 p 番目のボタンを提示した場合のものである。 N_q は、第 q 番目の層におけるセルの数であり、 $w_{ji}^{(q)}$ は第 $q-1$ 番目の層の第 i 番目のセルから第 q 番目の層の第 j 番目のセルへの重み係数、 $v_j^{(q)}$ は第 q 番目の層の第 j 番目のセルのバイアスである。本実施例では N_0 は512、 N_1 は64、 N_2 は1である。重み係数 $w_{ji}^{(q)}$ とバイアス $v_j^{(q)}$ は学習前にはランダムな小さな値に設

$$\delta_{p0}^{(2)} = (t_{p0} - o_{p0}^{(2)}) o_{p0}^{(2)} (1 - o_{p0}^{(2)}) \dots (13)$$

(4) 神経回路網の誤差逆伝搬学習処理(ステップS130)

第 q 番目の層の各セルの誤差 $\delta_{pj}^{(q)}$ が計算済みのとき、第 $(q-1)$ 層の各セルの誤差 $\delta_{pj}^{(q-1)}$ を次式

$$\delta_{pj}^{(q-1)} = o_{pj}^{(q-1)} (1 - o_{pj}^{(q-1)}) \sum_{k=0}^{N_q-1} \delta_{pk}^{(q)} w_{kj}^{(q)} \dots (14)$$

さらに誤差 $\delta_{pj}^{(q)}$ を用いて第 $(q-1)$ 層から第 q 層への重み係数 $w_{ji}^{(q)}$ の修正量 $\Delta p W_{ji}^{(q)}$ を次式(1

$$\Delta p W_{ji}^{(q)} = \eta \delta_{pj}^{(q)} o_{pi}^{(q-1)} \dots (15)$$

また、第 q 層の第 j 番目のセルのバイアス $v_j^{(q)}$ に対する修正量 $\Delta p W_{jv}^{(q)}$ も次式(16)で計算する。

$$\Delta p W_{jv}^{(q)} = \eta \delta_{pj}^{(q)} \dots (16)$$

ここで、 η は学習の速度を決定する定数である。これらの計算を層の番号 q を減じながら全ての出力層、中間層について実行し、全ての重み係数 $w_{ji}^{(q)}$ 及びバイアスに対する修正量 $\Delta p W_{ji}^{(q)}$ 及び $\Delta p W_{jv}^{(q)}$ を求め

$$W_{ji}^{(q)} = W_{ji}^{(q)} + \Delta p W_{ji}^{(q)} \dots (17)$$

以上の処理を $p=1$ として、提示信号 $s_{a0.9}(m)$ に対して行う。この時、教師信号 t_{10} としては0.9を与える。

【0015】(5) 学習終了判定処理(ステップS140)

学習処理の繰り返しによって重み係数 $w_{ji}^{(q)}$ が最適な値に近づくと、出力誤差 $\delta_{p0}^{(2)}$ が0に近づく。出力誤差 $\delta_{p0}^{(2)}$ が十分に小さな値 ε よりも小さな値になったかどうかを判定し、出力誤差 $\delta_{p0}^{(2)}$ が大きければ学習未終了であると判定し、処理ステップS100に戻る。

定しておく。数12の計算をすべての q に対して計算し、第2番目の出力層のセルの出力 $o_{p0}^{(2)}$ を得る。

【0013】(3) 神経回路網の出力誤差計算処理(ステップS120)

q 番目の層の第 j 番目のセルにおける第 p 番目の入力に対する誤差を $\delta_{pj}^{(q)}$ で表すと、第 q 番目の提示信号に対する教師信号を t_{pj} として、出力層における誤差 $\delta_{p0}^{(2)}$ を次式(13)で計算する。

(14)で計算しておく。

【数6】

5)で計算する。

【0014】

る。全ての修正量 $\Delta p W_{ji}^{(q)}$ 及び $\Delta p W_{jv}^{(q)}$ を計算した後、この修正量を用いて全ての重み係数 $w_{ji}^{(q)}$ 及びバイアス $v_j^{(q)}$ を次式(17)によって修正する。

出力誤差 $\delta_{p0}^{(2)}$ が小さければ、すべての学習処理を終了する。次に、処理ステップS100に戻った場合には、提示信号として前述の教師エポック点を中心としないうち320をとる。区間320の中心の教師エポック点からのずれはランダムとする。この場合の提示信号 $s_{a0.1}(m)$ を次式(18)でパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{2j}^{(0)}$ とする。

【数7】

$$o_{2j}^{(0)} = \frac{s_{a0.1}(j)}{C \sqrt{\frac{\sum_{m=0}^{511} s_{a0.1}(m)^2}{512}}} + 0.5 \quad \dots (18)$$

このときの教師信号 t_{20} は 0.1 とし、処理ステップ S110 以降は前述と同様の処理を行う。以上の学習処理を繰返し行うことにより、最終的に最適な重み係数が得られる。

【0016】次に、学習後の神経回路網を用いた基本周波数抽出処理について、図1(b)及び図5を参照しつつ説明する。なお、図5は基本周波数抽出処理における各信号の説明図であり、符号400は基本周波数抽出の対象となる入力信号、符号420は神経回路網に入力するある1つの入力信号の区間、符号421は区間420に対する神経回路網の出力、符号430は神経回路網に入力する次の1つの入力信号の区間、符号431は区間430に対する神経回路網の出力、符号440は基本周波数抽出処理によって得られる神経回路網の出力系列、及び符号450は出力系列から抽出されたエポック点を

$$o_{uj}^{(0)} = \frac{s_u(j)}{C \sqrt{\frac{\sum_{m=0}^{511} s_u(m)^2}{512}}} + 0.5 \quad \dots (19)$$

【0018】(B) 神経回路網の順方向伝搬処理 (ステップS160)

この順方向伝搬処理は、式(12)における p を u に置き換えて計算することにより、学習処理における順方向伝搬処理と同様に行う。この処理によって出力層のセルからは出力 $o_{u0}^{(2)}$ が得られる。さらに、入力信号を時刻 $u+1$ を中心とする時間区間からとり、同様な処理を行う。この様な処理を繰返し、時刻 u に対する出力 $o_{u0}^{(2)}$ の系列を得る。この出力系列の一例を図5の符号440として示す。入力信号を取り出した時間区間の中心と入力信号のエポック点とが一致するすると、出力系列

$$o_{v_{d-1}0}^{(2)} \leq o_{vd0}^{(2)} \geq o_{v_{d+1}0}^{(2)}, \quad P \leq o_{vd0}^{(2)}$$

... (20)

ここで、 P はピークを検出するための閾値であり、本実施例では定数 0.5 を用いる。 d は検出したピークに付与する番号である。

(D) 基本周波数算出処理 (ステップS180)

$$f_d = f_s / (v_d - v_{d-1})$$

本実施例は、信号波形のエポック点を自動的に検出する

示す。

【0017】(A) 神経回路網の信号入力処理 (ステップS150)

時間領域で離散的な基本周波数が未知の入力信号を $x(m)$ とし、本実施例では学習処理に用いた提示信号とは別の話者の母音音声を入力信号として用いることにする。ここで、現在着目している時刻 u を中心とした時間区間における入力信号を $x_u^{(m)}$ とする。この入力信号を神経回路網の入力として次式(19)によってパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{uj}^{(0)}$ とする。ここで、 $o_{uj}^{(q)}$ は q 番目の層における時刻 u を中心とした入力信号に対する j 番目のセルの出力を示し、 C は正規化のための正定数である。

【数8】

にはピークが生ずる。このピークを検出し、ピーク間の間隔を測定することによって入力信号の基本周波数を求めることができる。

【0019】(C) 神経回路網の出力ピーク検出処理 (ステップS170)

出力系列に対するピーク検出処理は、出力が次式(20)の条件を満たす離散的な時刻 v_d をエポック点の時刻として検出する。

【数9】

$$o_{v_{d-1}0}^{(2)} \leq o_{vd0}^{(2)} \geq o_{v_{d+1}0}^{(2)}, \quad P \leq o_{vd0}^{(2)}$$

... (20)

検出したエポック点の時刻 v_d の間隔とサンプリング周波数 f_s から d 番目のピッチにおける基本周波数 f_d を次式(21)により計算する。

$$\dots (21)$$

神経回路網を学習処理によって構成し、非線形処理によ

ってエポック点の時刻を強調して出力するので、従来の線形予測法による残差信号を用いた基本周波数抽出方法に比べて以下のような利点がある。

【0020】(1) 神経回路網の出力セルから得られる出力 $o_{u0}^{(2)}$ は、その値が0.0から1.0までの間であり、閾値 p は厳密な設定が不要である。

(2) 神経回路網の出力セルから得られる出力 $o_{u0}^{(2)}$ は、単純なパルス列となっており、ピーク位置の検出が容易であり、正確な基本周波数が検出しやすい。以上の利点を明確に示すために、実際のデータによって得られた各信号の例を図6(a), (b), (c)に示す。同図(a)は男性が発生した「ア」の一部である入力信号 $s(m)$ の波形図、同図(b)は線形予測分析によって得られた予測残差信号 $e(n)$ の波形図、及び同図(c)は神経回路網によって得られた出力系列 $o_{u0}^{(2)}$ の波形図である。なお、図中の縦線は人間の視察によって得られたエポック点の時刻であり、離散的な時刻 m, n, u の位置が互いに同一の時刻の部分を示している。

【0021】図6(b)で明らかなように、線形予測分析による残差信号波形 $e(n)$ は複雑であり、この信号波形からエポック点を正確に抽出して基本周波数を計算することは難しい。一方、神経回路網による出力系列 $o_{u0}^{(2)}$ は、学習処理によって生成された最適な非線形処理により単純なインパルス列の形状を示し、この出力系列からエポック点を正確に抽出して基本周波数を計算することは容易である。なお、本発明は、図示の実施例に限定されず、種々の変形が可能である。例えば、図4の提示信号300としては実施例に用いた以外の他の母音波形などの任意の周期性信号を与え、音響信号、画像信号等の一般的な信号波形に対してエポック点が検出できる神経回路網を構成することも可能である。

【0022】

【発明の効果】以上詳細に説明したように、本発明によれば、信号波形のエポック点を自動的に検出する神経回路網を学習処理によって構成し、その神経回路網に所定

の時間区間にわたって正規化された入力信号を入力する。さらに、入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークをエポック点として検出し、検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出するようにした。そのため、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出することが可能となる。また、提示信号及び入力信号を電力によって正規化すれば、正規化処理が簡単かつ的確に行える。

【図面の簡単な説明】

【図1】本発明の実施例を示す基本周波数抽出方法のフローチャートであり、同図(a)は学習処理のフローチャート及び同図(b)は基本周波数抽出処理のフローチャートである。

【図2】図2は従来の基本周波数抽出方法のフローチャートである。

【図3】本発明の実施例の基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。

【図4】本発明の実施例の学習処理における各信号の説明図である。

【図5】本発明の実施例の基本周波数抽出処理における各信号の説明図である。

【図6】本発明の効果を示す図であり、同図(a)は入力信号の波形図、同図(b)は従来の線形予測分析による波形図、及び同図(c)は神経回路網による出力系列の波形図である。

【符号の説明】

51	信号波形
52	セル
53	通信回路
54, 440	出力系列
300	提示信号
310, 450	エポック点
320, 330, 420, 430	時間区間
400	入力信号

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第2区分

【発行日】平成11年(1999)10月15日

【公開番号】特開平4-241400

【公開日】平成4年(1992)8月28日

【年通号数】公開特許公報4-2414

【出願番号】特願平3-3219

【国際特許分類第6版】

G10L 9/08

G06F 15/18

【FI】

G10L 9/08 B

G06F 15/18

【手続補正書】

【提出日】平成9年8月4日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正内容】

【書類名】明細書

【発明の名称】基本周波数抽出方法

【特許請求の範囲】

【請求項1】 所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、
所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、
前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、
該検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出することを特徴とする基本周波数抽出方法。

【請求項2】 請求項1記載の基本周波数抽出方法において、

前記提示信号及び前記入力信号は、電力によって正規化した基本周波数抽出方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声信号、音響信号、

$$s_n(m) = s(m+n)w(m) \quad \dots (1)$$

但し、 $0 \leq m \leq N-1$ 、(N ；所望の窓関数の大きさ)

(2) ステップS2の処理

所望の時刻 n における信号波形の自己相関関数 R

$n(k)$ を次式(2)により求める。

及び画像信号等の信号の基本周波数を抽出する基本周波数抽出方法に関するものである。

【0002】

【従来の技術】従来、この種の分野の技術としては、例えば図2のようなものがあった。以下、図2を参照しつつ従来の基本周波数抽出方法について説明する。図2は従来の基本周波数抽出方法の処理手順を示すフローチャートであり、各処理を順番にステップS1～S6で表す。まず、処理の対象となる信号波形に窓関数を乗じて所望の時間領域の信号を取り出した後(ステップS1)、信号波形の自己相関関数を求める(ステップS2)。続いて、ステップS2で求められた自己相関関数をもとに線形予測係数を求め(ステップS3)、さらに、この線形予測係数に基づき信号波形に対する線形予測を行って予測残差信号を求める(ステップS4)。その後、予測残差信号のピークを抽出してエポック点を検出し(ステップS5)、エポック点の時間間隔から基本周波数を計算する(ステップS6)。ここで、エポック点とは、信号波形の基本周波数に対応した時刻であり、音声信号の場合は、例えば声帯が最も収縮した発声時の時刻をいう。

【0003】次に、各処理の内容を説明する。

(1) ステップS1の処理

時間領域で離散的な信号波形を $s(m)$ とし(但し、 m は離散的な時刻)、適当な窓関数を $w(m)$ とする。今、処理対象となる信号波形のうち、基本周波数を求めたい所望の離散的な時刻を n とする。このとき、所望の時刻 n における信号波形 $s_n(m)$ は次式(1)で求める。

【数1】

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k) \quad \dots (2)$$

但し、 k : ラグ

【0004】(3) ステップS3の処理
線形予測係数を求める。ダービン(Durbin)の再帰法によれば、次式(3)～(7)によって線形予測係数
 $E^{(0)} = R(0)$

数 α_j を求めることができる。ここで、 j は第 j 次の予測係数であることを示し、 p は線形予測の次数で任意である。

$\dots (3)$

【数2】

$$k_i = (R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j)) / E^{(i-1)} \quad \dots (4)$$

但し、 $1 \leq i \leq p$

$$\alpha_j^{(i)} = k_i \quad \dots (5)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \quad \dots (6)$$

但し、 $1 \leq j \leq i-1$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad \dots (7)$$

計算の手順としては、まず、式(3)により $E^{(0)}$ を得る。次に、式(4)から式(7)までの計算を行い、 α

$j^{(i)}$ を $1 \leq i \leq p$ の範囲で順に再帰的に求める。そして、最後に、

$$\alpha_j = \alpha_j^{(p)} \quad \dots (8)$$

但し、 $1 \leq j \leq p$

として線形予測係数 α_j を得る。

れる。

【0005】(4) ステップS4の処理
求めた線形予測係数 α_j を用いて予測残差信号 $e(n)$ を求める。予測残差信号 $e(n)$ は次式(9)で計算さ

【数3】

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad \dots (9)$$

(5) ステップS5の処理
残差信号 $e(n)$ のピークをエポック点として求める。この処理はある一定の閾値 H を定め、残差信号 $e(n)$ がこの閾値 H を終えてピークを生じた離散的な時刻をエポック点とし、各エポック点の離散的な時刻を T_0, T

t_1, T_2, \dots とする。

(6) ステップS6の処理

最後にエポック点の間隔とサンプリング周波数 f_s から、 b 番目のピッチにおける基本周波数 f_b を次式(10)により計算する。

$$f_b = f_s / (T_b - T_{b-1}) \quad \dots (10)$$

抽出が困難である点について解決した基本周波数抽出方法を提供するものである。

【0006】

【発明が解決しようとする課題】しかしながら、上記の基本周波数抽出方法では、次のような課題があった。

① 予測残差信号 $e(n)$ の波高値の変化範囲は一定ではないため、エポック点検出のための閾値 H の設定によってはエポック点が正確に検出できない。

② 予測残差信号 $e(n)$ の波形は複雑であり、ピークが多数生ずる。そのため、どのピークがエポック点によるものが判別しにくく、正確な基本周波数の検出が困難である。

本発明は前記従来技術の持っていた課題として、正確なエポック点の検出が困難である点、正確な基本周波数の

【0007】

【課題を解決するための手段】本発明は、前記課題を解決するために、所定の時間区間にわたって正規化された学習用の提示信号を入力し、誤差逆伝搬法により前記時間区間の所定の位置と前記提示信号のエポック点との一致時のみに出力がピークとなるように学習処理が施された神経回路網を用い、所定の時間区間にわたって正規化された入力信号を前記神経回路網に入力し、前記入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークを前記エポック点として複数、検出し、該検出された各エポック点の時間間隔から前記入力信号の基

本周波数を算出するようにしたものである。また、前記提示信号及び前記入力信号は、電力によって正規化してもよい。

【0008】

【作用】本発明は、以上のように基本周波数抽出方法を構成したので、神経回路網に提示信号を入力させて誤差逆伝搬法により学習処理を施せば、神経回路網は、入力された提示信号の時間区間の所定の位置とエポック点との一致時のみに出力がピークとなるような信号波形を出力して、基本周波数を有する信号が入力されるとその信号のエポック点が自動的に検出されるようになる。このように学習がなされた神経回路網に基本周波数が未知である入力信号を入力し、その入力信号の時刻をわずかつ移動させて得られる神経回路網の出力系列のピークをエポック点として複数、検出した後、そのエポック点の時間間隔から入力信号の基本周波数を計算する。このように、非線形処理によってエポック点の時刻を強調して出力することにより、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出できる。したがって、前記課題を解決できるのである。

【0009】

【実施例】図1(a)、(b)は、本発明の実施例を示す基本周波数抽出方法の処理手順を示すフローチャートであり、同図(a)は学習処理の手順を示すフローチャート及び同図(b)は基本周波数抽出処理の手順を示すフローチャートである。また、図3は本発明の実施例を示す基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。図3において、この基本周波数抽出装置は、処理対象となる音声信号等の信号波形51をアナログ／デジタル変換してデジタルデータを求める図示しないアナログ／デジタル変換器を有し、その出力側には入力層を形成する複数の処理ユニット（以下、単にセルという）52が接続されている。さらに、通信回路53を介して入力層の各セル52が中間層のセル52に、中間層の各セル52が出力層のセル52にそれぞれ接続されている。これら各セル52は神経細胞に相当するものであり、それぞれマイクロプロセッサで構成されている。通信回路53は神経細胞間の接続ニューロンに対応するものであり、ローカルエリアネ

$$o_{ij}^{(0)} = \frac{S_{a0.9}(j)}{C \sqrt{\sum_{m=0}^{511} S_{a0.9}(m)^2}} + 0.5$$

(2) 神経回路網の順方向伝搬処理（ステップS110）

本実施例の神経回路網の構造は入力層を第0番目の層として、第1番目の層を中間層、第2番目の層を出力層と

ットワークで構成されている。さらに、出力層のセル52から出力される出力系列54をデジタルデータとして蓄積する図示しないメモリが設けられている。このようなハード構成により、非線形強調処理を行うパーセプトロン型の神経回路網が形成される。

【0010】この神経回路網の処理内容を以下に説明する。本発明の処理は、神経回路網の学習処理とその学習処理が終わった神経回路網による基本周波数の抽出処理とに分かれる。まず、神経回路網の学習処理について図1(a)及び図4を参照しつつ説明する。なお、図4は学習処理における各信号の説明図である。神経回路網の学習のために入力として加える提示信号300として、男性の発生した母音波形を12kHz、12ビットでサンプリングしたものをを用い、さらに、提示信号300に対して人間の視察により教師信号のピークを与えるエポック点の位置310（以下、教師エポック点という）を予め設定しておく。また、時間領域で離散的な信号波形s(m)とし、提示信号300を特にs_a(m)とする。

【0011】(1) 提示信号入力処理（ステップS100）

提示信号300を入力するための処理は次のようにして行われる。まず、提示信号300を教師エポック点を中心とした区間から取り出し、教師信号として0.9を定める。この場合の提示信号300の時間区間330の長さは神経回路網の入力層のセル数に等しいサンプル数とし、本実施例では512点のサンプルを提示信号300とする。この場合の提示信号300をs_{a0.9}(m)で表す。但し、0 ≤ m ≤ 511とする。そして、神経回路網の入力として、この提示信号s_{a0.9}(m)を次式(1)によってパワ正規化してオフセットを加え、入力層の各セルの出力o_{ij}⁽⁰⁾とする。なお、o_{pj}^(q)はq番目の層におけるp番目のパタンに対するj番目のセルの出力を表し、ここでは、提示信号s_{a0.9}(m)を1番目のパタンとし、入力層を0番目の層としている。Cは正規化のための正定数である。

【0012】

【数4】

$$\dots (11)$$

する3層構造であり、第q番目の層の出力は次式(12)で計算する。

【数5】

$$o_{pj}^{(q)} = 1 / (1 + \exp(-(\sum_{i=0}^{N_q-1} w_{ji}^{(q)} o_{pi}^{(q-1)} + v_j^{(q)}))) \quad \dots (12)$$

ここで、 $1 \leq q \leq 2$ である。式 (12) 中の $o_{pj}^{(q)}$ は第 q 番目の層における第 j 番目のセルの出力であり、第 p 番目のボタンを提示した場合のものである。 N_q は、第 q 番目の層におけるセルの数であり、 $w_{ji}^{(q)}$ は第 $q-1$ 番目の層の第 i 番目のセルから第 q 番目の層の第 j 番目のセルへの重み係数、 $v_j^{(q)}$ は第 q 番目の層の第 j 番目のセルのバイアスである。本実施例では N_0 は 512、 N_1 は 64、 N_2 は 1 である。重み係数 $w_{ji}^{(q)}$ とバイアス $v_j^{(q)}$ は学習前にはランダムな小さな値に設

$$\delta_{p0}^{(2)} = (t_{p0} - o_{p0}^{(2)}) o_{p0}^{(2)} (1 - o_{p0}^{(2)}) \quad \dots (13)$$

(4) 神経回路網の誤差逆伝搬学習処理 (ステップ S130)

第 q 番目の層の各セルの誤差 $\delta_{pj}^{(q)}$ が計算済みのもと、第 $(q-1)$ 層の各セルの誤差 $\delta_{pj}^{(q-1)}$ を次式

$$\delta_{pj}^{(q-1)} = o_{pj}^{(q-1)} (1 - o_{pj}^{(q-1)}) \sum_{k=0}^{N_q-1} \delta_{pk}^{(q)} w_{kj}^{(q)} \quad \dots (14)$$

さらに誤差 $\delta_{pj}^{(q)}$ を用いて第 $(q-1)$ 層から第 q 層への重み係数 $w_{ji}^{(q)}$ の修正量 $\Delta p W_{ji}^{(q)}$ を次式 (1

$$\Delta p W_{ji}^{(q)} = \eta \delta_{pj}^{(q)} o_{pi}^{(q-1)} \quad \dots (15)$$

また、第 q 層の第 j 番目のセルのバイアス $v_j^{(q)}$ に対する修正量 $\Delta p W_{jv}^{(q)}$ も次式 (16) で計算する。

$$\Delta p W_{jv}^{(q)} = \eta \delta_{pj}^{(q)} \quad \dots (16)$$

ここで、 η は学習の速度を決定する定数である。これらの計算を層の番号 q を減じながら全ての出力層、中間層について実行し、全ての重み係数 $w_{ji}^{(q)}$ 及びバイアスに対する修正量 $\Delta p W_{ji}^{(q)}$ 及び $\Delta p v_j^{(q)}$ を求め

$$W_{ji}^{(q)} = W_{ji}^{(q)} + \Delta p W_{ji}^{(q)} \quad \dots (17)$$

以上の処理を $p=1$ として、提示信号 $s_{a0.9}(m)$ に対して行う。この時、教師信号 t_{i0} としては 0.9 を与える。

【0015】(5) 学習終了判定処理 (ステップ S140)

学習処理の繰り返しによって重み係数 $w_{ji}^{(q)}$ が最適な値に近づくと、出力誤差 $\delta_{p0}^{(2)}$ が 0 に近づく。出力誤差 $\delta_{p0}^{(2)}$ が十分に小さな値 ϵ よりも小さな値になったかどうかを判定し、出力誤差 $\delta_{p0}^{(2)}$ が大きければ学習未終了であると判定し、処理ステップ S100 に戻る。

定しておく。数 12 の計算をすべての q に対して計算し、第 2 番目の出力層のセルの出力 $o_{p0}^{(2)}$ を得る。

【0013】(3) 神経回路網の出力誤差計算処理 (ステップ S120)

q 番目の層の第 j 番目のセルにおける第 p 番目の入力に対する誤差を $\delta_{pj}^{(q)}$ で表すと、第 q 番目の提示信号に対する教師信号を t_{pj} として、出力層における誤差 $\delta_{p0}^{(2)}$ を次式 (13) で計算する。

(14) で計算しておく。
【数 6】

5) で計算する。

【0014】

る。全ての修正量 $\Delta p W_{ji}^{(q)}$ 及び $\Delta p v_j^{(q)}$ を計算した後、この修正量を用いて全ての重み係数 $w_{ji}^{(q)}$ 及びバイアス $v_j^{(q)}$ を次式 (17) によって修正する。

出力誤差 $\delta_{p0}^{(2)}$ が小さければ、すべての学習処理を終了する。次に、処理ステップ S100 に戻った場合には、提示信号として前述の教師エポック点を中心としない区間 320 をとる。区間 320 の中心の教師エポック点からのずれはランダムとする。この場合の提示信号 $s_{a0.1}(m)$ を次式 (18) でパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{2j}^{(0)}$ とする。

【数 7】

$$o_{2j}^{(0)} = \frac{s_{a0.1(j)}}{C \sqrt{\frac{\sum_{m=0}^{511} s_{a0.1(m)}^2}{512}}} + 0.5 \quad \dots (18)$$

このときの教師信号 t_{20} は 0.1 とし、処理ステップ S110 以降は前述と同様の処理を行う。以上の学習処理を繰返し行うことにより、最終的に最適な重み係数が得られる。

【0016】次に、学習後の神経回路網を用いた基本周波数抽出処理について、図1(b)及び図5を参照しつつ説明する。なお、図5は基本周波数抽出処理における各信号の説明図であり、符号400は基本周波数抽出の対象となる入力信号、符号420は神経回路網に入力するある1つの入力信号の区間、符号421は区間420に対する神経回路網の出力、符号430は神経回路網に入力する次の1つの入力信号の区間、符号431は区間430に対する神経回路網の出力、符号440は基本周波数抽出処理によって得られる神経回路網の出力系列、及び符号450は出力系列から抽出されたエポック点を

$$o_{uj}^{(0)} = \frac{s_d(j)}{C \sqrt{\frac{\sum_{m=0}^{511} s_d(m)^2}{512}}} + 0.5 \quad \dots (19)$$

【0018】(B) 神経回路網の順方向伝搬処理 (ステップS160)

この順方向伝搬処理は、式(12)における p を u に置き換えて計算することにより、学習処理における順方向伝搬処理と同様に行う。この処理によって出力層のセルからは出力 $o_{u0}^{(2)}$ が得られる。さらに、入力信号を時刻 $u+1$ を中心とする時間区間からとり、同様な処理を行う。この様な処理を繰返し、時刻 u に対する出力 $o_{u0}^{(2)}$ の系列を得る。この出力系列の一例を図5の符号440として示す。入力信号を取り出した時間区間の中心と入力信号のエポック点とが一致するすると、出力系列

$$o_{v_{d-1}0}^{(2)} \leq o_{vd0}^{(2)} \geq o_{v_{d+1}0}^{(2)}, P \leq o_{vd0}^{(2)} \quad \dots (20)$$

ここで、 P はピークを検出するための閾値であり、本実施例では定数 0.5 を用いる。 d は検出したピークに付与する番号である。

(D) 基本周波数算出処理 (ステップS180)

$$f_d = f_s / (v_d - v_{d-1}) \quad \dots (21)$$

本実施例は、信号波形のエポック点を自動的に検出する

示す。

【0017】(A) 神経回路網の信号入力処理 (ステップS150)

時間領域で離散的な基本周波数が未知の入力信号を $x(m)$ とし、本実施例では学習処理に用いた提示信号とは別の話者の母音音声を入力信号として用いることにする。ここで、現在着目している時刻 u を中心とした時間区間における入力信号を $x_u(m)$ とする。この入力信号を神経回路網の入力として次式(19)によってパワ正規化してオフセットを加え、入力層の各セルの出力 $o_{uj}^{(0)}$ とする。ここで、 $o_{uj}^{(q)}$ は q 番目の層における時刻 u を中心とした入力信号に対する j 番目のセルの出力を示し、 C は正規化のための正定数である。

【数8】

にはピークが生ずる。このピークを検出し、ピーク間の間隔を測定することによって入力信号の基本周波数を求めることができる。

【0019】(C) 神経回路網の出力ピーク検出処理 (ステップS170)

出力系列に対するピーク検出処理は、出力が次式(20)の条件を満たす離散的な時刻 v_d をエポック点の時刻として検出する。

【数9】

検出したエポック点の時刻 v_d の間隔とサンプリング周波数 f_s から、 d 番目のピッチにおける基本周波数 f_d を次式(21)により計算する。

神経回路網を学習処理によって構成し、非線形処理によ

ってエポック点の時刻を強調して出力するので、従来の線形予測法による残差信号を用いた基本周波数抽出方法に比べて以下のような利点がある。

【0020】(1) 神経回路網の出力セルから得られる出力 $ou_0^{(2)}$ は、その値が0.0から1.0までの間であり、閾値 p は厳密な設定が不要である。

(2) 神経回路網の出力セルから得られる出力 $ou_0^{(2)}$ は、単純なパルス列となっており、ピーク位置の検出が容易であり、正確な基本周波数が検出しやすい。以上の利点を明確に示すために、実際のデータによって得られた各信号の例を図6(a), (b), (c)に示す。同図(a)は男性が発生した「ア」の一部である入力信号 $s(m)$ の波形図、同図(b)は線形予測分析によって得られた予測残差信号 $e(n)$ の波形図、及び同図(c)は神経回路網によって得られた出力系列 $ou_0^{(2)}$ の波形図である。なお、図中の縦線は人間の視察によって得られたエポック点の時刻であり、離散的な時刻 m, n, u の位置が互いに同一の時刻の部分を示している。

【0021】図6(b)で明らかなように、線形予測分析による残差信号波形 $e(n)$ は複雑であり、この信号波形からエポック点を正確に抽出して基本周波数を計算することは難しい。一方、神経回路網による出力系列 $ou_0^{(2)}$ は、学習処理によって生成された最適な非線形処理により単純なインパルス列の形状を示し、この出力系列からエポック点を正確に抽出して基本周波数を計算することは容易である。なお、本発明は、図示の実施例に限定されず、種々の変形が可能である。例えば、図4の提示信号300としては実施例に用いた以外の他の母音波形などの任意の周期性信号を与え、音響信号、画像信号等の一般的な信号波形に対してエポック点が検出できる神経回路網を構成することも可能である。

【0022】

【発明の効果】以上詳細に説明したように、本発明によれば、信号波形のエポック点を自動的に検出する神経回路網を学習処理によって構成し、その神経回路網に所定

の時間区間にわたって正規化された入力信号を入力する。さらに、入力信号の時刻を移動させて得られる前記神経回路網の出力系列のピークをエポック点として検出し、検出された各エポック点の時間間隔から前記入力信号の基本周波数を算出するようにした。そのため、基本周波数に相当するエポック点の時刻を正確に検出でき、正確な基本周波数を抽出することが可能となる。また、提示信号及び入力信号を電力によって正規化すれば、正規化処理が簡単かつ的確に行える。

【図面の簡単な説明】

【図1】本発明の実施例を示す基本周波数抽出方法のフローチャートであり、同図(a)は学習処理のフローチャート及び同図(b)は基本周波数抽出処理のフローチャートである。

【図2】図2は従来の基本周波数抽出方法のフローチャートである。

【図3】本発明の実施例の基本周波数抽出方法を実施するための基本周波数抽出装置の概略の構成図である。

【図4】本発明の実施例の学習処理における各信号の説明図である。

【図5】本発明の実施例の基本周波数抽出処理における各信号の説明図である。

【図6】本発明の効果を示す図であり、同図(a)は入力信号の波形図、同図(b)は従来の線形予測分析による波形図、及び同図(c)は神経回路網による出力系列の波形図である。

【符号の説明】

51	信号波形
52	セル
53	通信回路
54, 440	出力系列
300	提示信号
310, 450	エポック点
320, 330, 420, 430	時間区間
400	入力信号

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.